Mathematical models of haploinsufficiency

Indrani Bose and Rajesh Karmakar

4th February 2008

Department of Physics
Bose Institute
93/1, Acharya Prafulla Chandra Road, Kolkata-700 009, India

Abstract

We study simple mathematical models of gene expression to explore the possible origins of haploinsufficiency (HI). In a diploid organism, each gene exists in two copies and when one of these is mutated, the amount of proteins synthesized is reduced and may fall below a threshold level for the onset of some desired activity. This can give rise to HI, a manifestation of which is in the form of a disease. We consider both deterministic and stochastic models of gene expression and suggest possible scenarios for the occurrence of HI in the two cases. In the stochastic case, random fluctuations around the mean protein level give rise to a finite probability that the protein level falls below a threshold. Increased gene copy number and faster gene expression kinetics reduce the variance around the mean protein level. The difference between slow and fast gene expression kinetics, as regards response to a signaling gradient, is further pointed out. The majority of results reported in the paper are derived analytically.

PACS: 05.10.Gg, 82.30.-k, 87.10.+e, 87.15.Aa

I. Introduction

Complex multicellular organisms are in general diploids, i.e., each cell in an organism contains two copies of the full set of genes in contrast to haploids in which each cell contains a single copy of the genome. Genes provide the blueprint for the synthesis of proteins which perform essential functions in cells. If one copy of a gene is mutated, there is approximately a 50% reduction in the level of proteins synthesized. In many cases this does not lead to observable changes and normalcy is retained. A common interpretation of haploinsufficiency (HI) is that it occurs when half normal levels of

proteins are insufficient for completing particular tasks, leading to specific types of diseases. More generally, HI may occur when the level of proteins synthesized falls below a critical level for the onset of some desired activity.

There is presently an extensive literature on the genetic and biomedical aspects of HI [1, 2, 3] but mathematical models exploring the origin of HI and the issues related to it are practically non-existent. It is by now well-accepted that stochastic processes have considerable effect on patterns of gene expression in cells [4, 5, 6, 7]. Cook et al [3] have studied the role of stochastic gene expression in HI by constructing a minimal model of gene expression and using numerical techniques to simulate the model. Their major finding is that when one of the two genes in a diploid organism is inactivated due to mutations, there is an increased susceptibility to stochastic initiations and interruptions of gene expression. As a result, the number of proteins produced may transiently fall below the desired level giving rise to HI. Both increased gene copy number and faster gene expression kinetics reduce expression noise, thus enhancing the possibility of a stable outcome.

A large number of diseases are caused by mutations in genes encoding proteins called transcription factors (TF). More than 30 different human maladies have been attributed to TF HI [1]. TFs regulate gene expression by binding at the promoter region of the gene to be expressed. Cooperative interactions among the TFs favour the formation of bound TF complexes (oligomers). The TFs interact at only one site or at multiple sites of the promoter. A simple mathematical model has been proposed to explore HI in systems involving cooperative assembly of TFs [2]. Such multimeric complexes are essential for initiation of gene expression in many eukaryotic systems. The model explores the relationship of fractional oligomerization Y with the free ($[S]$) as well as total concentrations of TFs ($[S_0]$). The TFs oligomerise to form a bound complex. The curves Y versus $[S]$ and $[S_0]$ have sigmoidal shapes. Due to the characteristic S shape of a sigmoid, a small change in the TF concentration around the inflection point (the point at which the tangent to the curve has the maximum slope) gives rise to a significant change in the magnitude of Y. Thus, if there are two TF-encoding genes and one of these becomes silent, the level of TFs produced may fall below the inflection point of the sigmoid and consequently the magnitude of Y, the fractional oligomerization, is considerably decreased. This results in reduced expression from the target gene, giving rise to TF HI if the amount of proteins synthesized falls below a threshold level.

In Section 2 of this paper, we extend the minimal model of Cook et al [3] to investigate the influence of bound complexes of TFs on the initiation of gene expression. In Section 3, we study the stochastic version of the minimal model and its extensions to elucidate the role of stochasticity in HI. We derive analytical expressions for the quantities determined numerically by Cook et al [3].

2. Deterministic Model

The model is an extension of the minimal model of gene expression studied by Cook et al [3]. A brief description of the model is as follows. A gene can be in two possible states: inactive (G) and active (G^*). Random transtions occur between the states G and G^* according to the first order reaction kinetics

$$G \underset{k_d}{\overset{k_a}{\rightleftharpoons}} G^\star \xrightarrow{j_p} p \xrightarrow{k_p} \Phi \tag{1}$$

where k_a and k_d are the activation and deactivation rate constants. The corresponding half-times are $T_a = \frac{log2}{k_a}$ and $T_d = \frac{log2}{k_d}$ respectively. In the active state G^*, the gene synthesizes a protein (p) with the rate constant j_p. The protein product degrades with a rate constant k_p and the associated half-time is T_p. The protein degradation product is represented as Φ. We now assume that activation to the state G^* is brought about by an inducing stimulus S, e.g., TFs. The reaction scheme in the presence of the stimulus is given by

$$G + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} GS \underset{k_d}{\overset{k_a}{\rightleftharpoons}} G^\star \xrightarrow{j_p} p \xrightarrow{k_p} \Phi \tag{2}$$

where GS represents the bound complex of G and S from which transition to the active state G^* occurs. If n_G is the total concentration of genes then

$$n_G = [G] + [GS] + [G^*] \tag{3}$$

where $[G]$, $[GS]$ and $[G^*]$ denote the concentrations of genes in the states G, GS and G^* respectively. In the steady state, we have

$$\frac{d[G]}{dt} = k_2[GS] - k_1[G][S] = 0$$

so that

$$\frac{[G][S]}{K_s} = [GS] \tag{4}$$

where $K_s = \frac{k_2}{k_1}$ is the equilibrium dissociation constant. From (3) and (4), we get

$$[GS] = \frac{n_G[S]/K_s}{1 + [S]/K_s} - [G^*]\frac{[S]/K_s}{1 + [S]/K_s} \tag{5}$$

Also, in the steady state,

$$\frac{d[G^*]}{dt} = k_a[GS] - k_d[G^*] = 0 \tag{6}$$

From (5) and (6), the expression for $[G^*]$ in the steady state is given by

$$[G^*] = \frac{n_G k_a \frac{[S]/K_s}{1+[S]/K_s}}{k_a \frac{[S]/K_s}{1+[S]/K_s} + k_d} \tag{7}$$

The reaction scheme in (1) leads to the expression

$$[G^*] = \frac{n_G k_a}{k_a + k_d} \tag{8}$$

in the steady state. Expression (7) and (8) are equivalent on defining effective activation and deactivation rate constants:

$$k_a^{'} = k_a \frac{[S]/K_s}{1 + [S]/K_s},$$

$$k_d^{'} = k_d \tag{9}$$

We now assume the inducing stimulus to be TFs. In the simplest approximation, n individual TFs oligomerise to produce an active complex S_n according to the reaction scheme

$$n\,S \quad \overset{K}{\rightleftharpoons} \quad S_n \tag{10}$$

The n TFs interact all at once to give rise to the bound complex $[S_n]$, i.e., we ignore the formation of dimers, tetramers, ... etc. Let $[S_0]$ be the initial concentration of TFs. Then

$$[S_0] = [S] + n\,[S_n] \tag{11}$$

where $[S]$ and $[S_n]$ are the concentrations of free TFs and the bound TF-complex respectively. The global equilibrium constant K is given by

$$K = \frac{[S_n]}{[S]^n} = \frac{[S_n]}{([S_0] - n\,[S_n])^n} \tag{12}$$

The fractional oligomerization is defined as [2]

$$Y = \frac{n\,[S_n]}{[S_0]} = \frac{n\,[S_n]}{[S] + n\,[S_n]} \tag{13}$$

Using (11) and (12), Y can further be written as

$$Y = \frac{[S]^n}{\frac{[S]}{K\,n} + [S]^n} \tag{14}$$

and

$$Y = \frac{([S_0] - n\,[S_n])^n}{\frac{([S_0] - n\,[S_n])}{K\,n} + ([S_0] - n\,[S_n])^n} \tag{15}$$

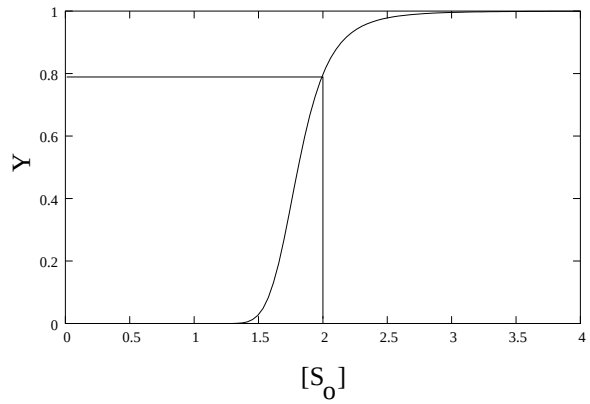

Figure 1: Fractional oligomerization Y versus $[S_0]$ (Eq. (15)) for $n = 6$, $K = 2$ and $[S_n] = 0.2$

Fig.1 shows the curve, fractional oligomerization Y versus $[S_0]$ (Eq. (15)) for $n = 6, K = 2$ and $[S_n] = 0.2$. The curve has the well-known sigmoidal shape.

We now replace $[S]$ by $[S_n]$ in the reaction scheme described by Eq. (2), i.e., we assume that the TF-oligomer S_n binds to a gene in the inactive state G to give rise to the bound complex GS_n. Transition to the active state G^* occurs from the intermediate state GS_n. The concentration $[G^*]$ in the steady state is obtained from (7) by replacing $[S]$ by $[S_n]$ where $[S_n] = K\,[S]^n$ (Eq. (12)). One finally obtains

$$[G^*] = \frac{n_G\, k_a\, \frac{K\,[S]^n/K_s}{1+K\,[S]^n/K_s}}{k_a\, \frac{K\,[S]^n/K_s}{1+K\,[S]^n/K_s} + k_d} \tag{16}$$

The concentration of proteins in the steady state is given by

$$[p] = \frac{j_p}{k_p}\,[G^*] \tag{17}$$

From (14), $[S]^n$ can be written as

$$[S]^n = \frac{Y}{1-Y}\,\frac{[S]}{K\,n} \tag{18}$$

From (16) and (17) and for $k_a = k_d$, we get

$$[p] = n_G\,\frac{j_p}{k_p}\,\frac{a\,Y}{1+2aY-Y} \tag{19}$$

where $a = \frac{[S]}{n\,K_s}$.

Fig. 2 shows the protein concentration $[p]$ versus fractional oligomerization Y for $k_a = k_d$ (Eq.(19)), $n_G = 2$, $j_p = 0.5\,k_p$, $K_s = 1.2$ and $n = 6$. From Eqs. (11), (16), and (17)

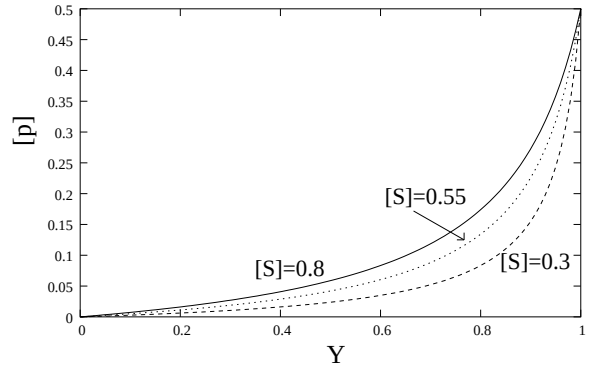

Figure 2: Protein concentration $[p]$ versus fractional oligomerization Y for $k_a = k_d$ (Eq. (19)), $n_G = 2$, $j_p = 0.5\, k_p$, $K_s = 1.2$ and $n = 6$.

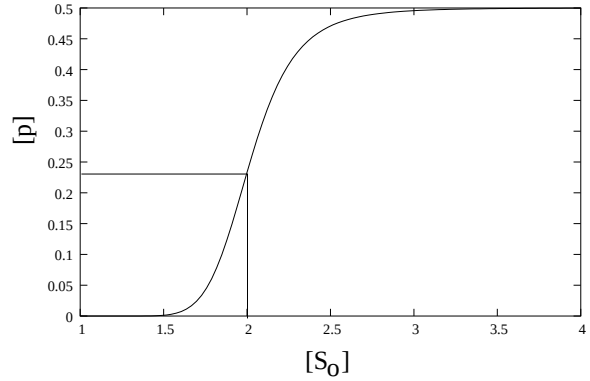

Figure 3: Protein concentration $[p]$ versus $[S_0]$ (Eq. (20))for $n = 6$, $K = 2$, $K_s = 1.2$ and $[S_n] = 0.2$

and for $k_a = k_d$, the concentration of protein $[p]$, as a function of the total concentration $[S_0]$ of TFs, can be written as

$$[p] = n_G \, \frac{j_p}{k_p} \, \frac{([S_0] - n\,[S_n])^n}{2\,([S_0] - n\,[S_n])^n + K_s/K} \tag{20}$$

Fig. 3 shows the plot $[p]$ versus $[S_0]$ (Eq. (20)) for $n = 6$, $K = 2$, $K_s = 1.2$ and $[S_n] = 0.2$. Figs. 1, 2 and 3 provide a possible explanation for the origin of HI. Suppose two TF-encoding genes produce TFs of total concentration $[S_0] = 4$. If one of the genes is inactivated due to mutations, the total concentration $[S_0]$ falls to the value 2. For the parameter values corresponding to Fig. 1, the fractional oligomerization Y has the value 0.797. The TFs form a bound complex S_n ($n = 6$) which then activates the gene synthesizing the protein P. The concentration of proteins $[p]$ corresponding to $Y = 0.797$ and for $[S] = 0.8$ is given by $[p] = 0.233$ (Fig. 2). The same value $[p]$ is obtained from Fig. 3 with $[S_0] = 2$. If both the encoding genes are active, the values of $[S_0]$, Y and P are $[S_0] = 4$, $Y = 1.0$ and $[p] = 0.5$ respectively. Thus, for one gene, the protein level is reduced by more than half. If the level falls below a threshold, the

amount of proteins synthesized is not sufficient for the execution of a particular task. This gives rise to HI, the manifestation of which is in the form of a disease.

3. Stochastic Approach

Let us first consider the model described in Section 2 in the absence of an inducing stimulus. Let n_{tot} be the total no of genes and n_0 , n_1 ($n_{tot} = n_0 + n_1$), the number of genes in the inactive (G) and active (G^*) states respectively. In the stochastic model, a gene makes random transitions between the inactive and active states with k_a and k_d being the activation and deactivation rate constants. In the active state, protein production and degradation occur with the rate constants j_p and k_p respectively. Let $p(n_1, n_2, t)$ be the probability that at time t, n_1 genes are in the active state G^* and the number of protein molecules is n_2. The rate of change of the probability with respect to time is given by the Master Equation

$$\begin{aligned}\frac{\partial p(n_1,n_2,t)}{\partial t} = \; & k_a[(n_{tot}-n_1+1)p(n_1-1,n_2,t)-(n_{tot}-n_1)p(n_1,n_2,t)] \\ & +k_d[(n_1+1)p(n_1+1,n_2,t)-n_1p(n_1,n_2,t)] \\ & +j_p[n_1p(n_1,n_2-1,t)-n_1p(n_1,n_2,t)] \\ & +k_p[(n_2+1)p(n_1,n_2+1,t)-n_2p(n_1,n_2,t)]\end{aligned} \tag{21}$$

For each rate constant, there is a gain term which adds to the probability and a loss term which subtracts from the probability.

We now use the standard approach in the theory of stochastic processes [8] to determine the average number of activated genes $< n_1 >$ and proteins $< n_2 >$ in the steady state and the variances thereof. Define the generating function

$$F(z_1,\, z_2,\, t) = \sum_{n_1,n_2} z_1^{n_1}\, z_2^{n_2}\, p(n_1,\, n_2, t) \tag{22}$$

In terms of the generating function, the Master equation (21) becomes

$$\frac{\partial F}{\partial t} = k_a n_{tot}(z_1-1)F - k_a(z_1-1)z_1\frac{\partial F}{\partial z_1} - k_d(z_1-1)\frac{\partial F}{\partial z_1} + j_p(z_2-1)z_1\frac{\partial F}{\partial z_1} - k_p(z_2-1)\frac{\partial F}{\partial z_2} \tag{23}$$

In the steady state $\frac{\partial F}{\partial t} = 0$. The following properties of the generating function are used in subsequent calculations:

$$F\mid_{z_1=1,z_2=1} = 1 \tag{24}$$

$$< n_1 > = \frac{\partial F}{\partial z_1}\mid_{z_1=1,z_2=1}, \quad < n_2 > = \frac{\partial F}{\partial z_2}\mid_{z_1=1,z_2=1} \tag{25}$$

where $< n_1 >$ is the mean number of active genes, i.e., genes in the state G^* and $< n_2 >$ is the same for proteins. Furthermore,

$$\begin{aligned}\frac{\partial^2 F}{\partial z_1^2}\mid_{z_1=1,z_2=1} &= < n_1^2 > - < n_1 > \\ \frac{\partial^2 F}{\partial z_2^2}\mid_{z_1=1,z_2=1} &= < n_2^2 > - < n_2 >\end{aligned} \tag{26}$$

Hence the variances around the mean levels are given by

$$\begin{aligned} Var_{n_1} &=< n_1^2 > - < n_1 >^2 = \frac{\partial^2 F}{\partial z_1^2} \mid_{z_1=1,z_2=1} + < n_1 > - < n_1 >^2 \\ Var_{n_2} &=< n_2^2 > - < n_2 >^2 = \frac{\partial^2 F}{\partial z_2^2} \mid_{z_1=1,z_2=1} + < n_2 > - < n_2 >^2 \end{aligned} \tag{27}$$

Successive differentiation of Eq. (23) ($\frac{\partial F}{\partial t} = 0.$) with respect to z_1 and z_2 gives rise to linear equations for successively higher moments. The equations may be solved to obtain, in particular, the mean and the variance. For example, differentiating Eq. (23) with respect to z_1 and z_2 and then putting $z_1,\ z_2 = 1$, one obtains expressions for the mean.
The mean and variance are given by

$$< n_1 >= \frac{n_{tot}\, k_a}{k_a + k_d} \tag{28}$$

$$Var_{n_1} =< n_1 > \frac{k_d}{k_a + k_d} \tag{29}$$

$$< n_2 >=< p >=< n_1 > \frac{j_p}{k_p} = \frac{j_p}{k_p} \frac{n_{tot}\, k_a}{k_a + k_d} \tag{30}$$

$$Var_{n_2} =< n_1 > \frac{j_p}{k_p} [1 + \frac{j_p\, k_d}{(k_a + k_d)(k_a + k_d + k_p)}] \tag{31}$$

As in Ref. 3, temporal quantities are scaled relative to the product half-life $T_p = \frac{log2}{k_p}$.
Let $T_a = \frac{log2}{k_a}$ and $T_d = \frac{log2}{k_d}$ be the times for half-maximal gene activation and deactivation respectively. The times T_a and T_d are scaled relative to T_p. Some of the results obtained in Ref. 3, using numerical simulation techniques, can readily be derived from the analytical expressions in (28)-(31). Stochasticity introduces random fluctuations around the mean protein level and variance gives a measure of the spread. Let $T_a = T_d = T_p/4$, i.e., $k_a = k_d = \alpha\, k_p$ with $\alpha > 0$. As α increases, one has faster expression kinetics and from (31) it is easy to verify that variance is reduced, i.e., the expression noise is less. The mean product level (Eq. (30)) is, however, independent of α. With increase in j_p, i.e., the protein synthesis rate, the variance increases. Let us now consider the case when the net expression rate of n_{tot} genes is distributed to one single gene so that the mean protein level remains the same. From (30)

$$< n_2 >= \frac{j_p}{k_p} \frac{n_{tot}}{2} = \frac{j_p^{'}}{k_p} \frac{1}{2} \tag{32}$$

where $j_p^{'} = j_p\, n_{tot}$ is the expression rate when only one gene is considered. Since $j_p^{'} > j_p$, the gene copy number is reduced from n_{tot} to 1. Similarly, when the net expression rate of n_{tot} genes is distributed to a larger number genes, say, from two to four, the variance is reduced.When one of two genes is inactivated due to mutations, the average protein level in the steady state is reduced by 50%. This may still be higher

than the threshold level required for protein activity. Due to the variance around the mean level, the number of proteins may transiently fall below the threshold giving rise to HI. The occurrence of HI further becomes more probable for slower expression kinetics as then the variance is increased in magnitude. For stochastic gene expression in the presence of an inducing stimulus, say, TF's, we use the effective model with the activation/deactivation rate constants given in Eq. (9). The expressions for the mean and the variance are the same as in Eqs. (28)-(31) but with k_a, k_d replaced by $k_a^{'}$ and $k_d^{'}$ respectively.

We now derive expressions for the probability distributions of protein levels in the steady state. To do this, we consider a simpler stochastic model in which the only stochasticity arises from the random transitions of a gene between the inactive and active states. In each state of the gene, the concentration of proteins evolves deterministically according to the equation

$$\frac{dx}{dt} = \frac{j_p}{X_{max}}\, z - k_p x = f(x,z) \tag{33}$$

where $z = 1\ (0)$ when the gene is in the active (inactive) state and $x = \frac{X}{X_{max}}$, X and X_{max} being the protein concentration at time t and the maximum protein concentration respectively. We note that $X_{max} = \frac{j_p}{k_p}$. Let $p_j(x,t)\ (j = 0, 1)$ be the probability density function when $z = j$. The total probability density function is

$$p(x,t) = p_0(x,t) + p_1(x,t) \tag{34}$$

The rate of change of probability density is given by

$$\frac{\partial p_j(x,t)}{\partial t} = -\frac{\partial}{\partial x}[f(x,j)p_j(x,t)] + \sum_{k\neq j}[W_{kj}\, p_k(x,t) - W_{jk}\, p_j(x,t)] \tag{35}$$

where W_{kj} is the transition rate from the state k to the state j and W_{jk} is the same for the reverse transition. The first term in Eq. (35) is the so called "transport" term representing the net flow of the probability density. The second term represents the gain/loss in the probability density due to random transitions between the state j and other accessible states. In the present case, Eq. (35) gives rise to the following two equations:

$$\frac{\partial p_0(x,t)}{\partial t} = -\frac{\partial}{\partial x}(-k_p x\, p_0(x,t)) + k_d\, p_1(x,t) - k_a\, p_0(x,t) \tag{36}$$

$$\frac{\partial p_1(x,t)}{\partial t} = -\frac{\partial}{\partial x}\{(\frac{j_p}{X_{max}} - k_p\, x)p_1(x,t)\} + k_a\, p_0(x,t) - k_d\, p_1(x,t) \tag{37}$$

The Master equation (Eq. (21)) provides a full stochastic description of all the processes associated with gene expression, namely, gene activation and deactivation, protein synthesis and degradation. The only stochastic events considered by Cook et al [3] are those related to gene activation and deactivation. In their model, protein synthesis

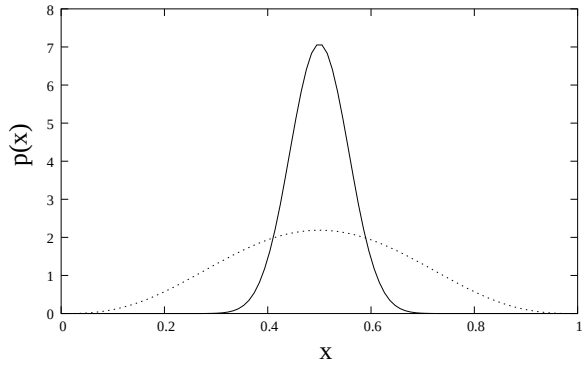

Figure 4: $p(x)$ versus x for slow (dotted curve) and fast (solid) gene expression kinetics

from the active gene and protein degradation occur in a deterministic manner. Eqs. (36) - (37) describe the scenario studied by Cook et al. The steady state distribution in this case is given by

$$p(x) = C\, x^{(\frac{k_a}{k_p}-1)}(1-x)^{(\frac{k_d}{k_p}-1)} \tag{38}$$

where C the normalization constant is given by the inverse of a beta function.

$$C = \frac{1}{B(\frac{k_a}{k_p}, \frac{k_d}{k_p})} \tag{39}$$

Since the probability density function is known, the mean protein level and its variance can be calculated in a straightforward manner. The mean protein level is identical to that obtained from the Master equation (Eq. (21)) whereas the variance is underestimated as stochasticity is taken into account only at the levels of gene activation and deactivation. Fig. 4 shows the plot of $p(x)$ versus x for slow ($T_a = T_d = T_p/4$) and fast ($T_a = T_d = T_p/40$) gene expression kinetics. In the latter case, the distribution is significantly narrower, i.e., faster kinetics lead to a reduction in the variance. The same conclusion is reached from the Master equation approach. From the full width at half maximum of the broader distribution, one finds that x ranges from 0.25 to 0.75, 0.5 being the mean value. Thus, it is probable that the protein level falls below the threshold for desired activity giving rise to HI. Let x_{thr} (< 1) be the threshold value of x. The probability that x is greater than x_{thr} is

$$p(x > x_{thr}) = 1 - \frac{\int_0^{x_{thr}} x^{(\frac{k_a}{k_p}-1)}(1-x)^{(\frac{k_d}{k_p}-1)}dx}{\int_0^1 x^{(\frac{k_a}{k_p}-1)}(1-x)^{(\frac{k_d}{k_p}-1)}dx} \tag{40}$$

$$= 1 - \frac{k_p\, x_{thr}^{\frac{k_a}{k_p}}\, {}_2F_1[1-\frac{k_d}{k_p}, \frac{k_a}{k_p}, 1+\frac{k_a}{k_p}, X_{thr}]}{k_a\, B(\frac{k_a}{k_p}, \frac{k_d}{k_p})} \tag{41}$$

where ${}_2F_1(a, b, c; z)$ is the hypergeometric function.

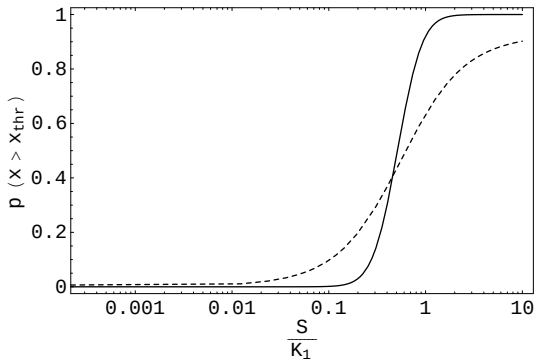

Figure 5: $p(x > x_{thr})$ versus S/K_s in a semi-logarithm plot for $n = 1$. The solid (dotted) curve corresponds to fast (slow) kinetics.

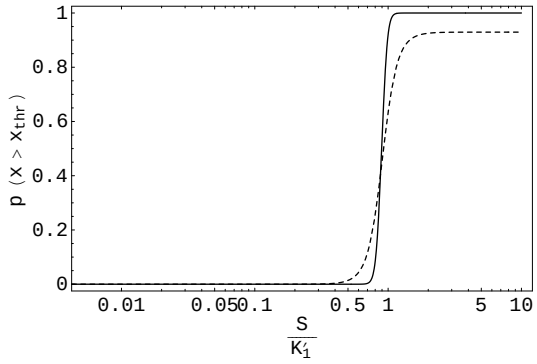

Figure 6: $p(x > x_{thr})$ versus $S/K_s^{'}$ $[(K_s^{'})^n = K_s/K]$ in a semi-logarithm plot for $n = 6$. The solid (dotted) curve corresponds to fast (slow) kinetics.

Let x_{thr} be 0.25. The probability $p(x > x_{thr})$ is computed using Mathematica for both slow and fast gene expression kinetics. The values of $p(x > x_{thr})$ in the slow and fast cases are 0.9294 and 0.9999 respectively. Since in the latter case, the probability that the protein level exceeds the threshold is higher, the chance of HI occurrence is correspondingly lower.

In the presence of an inducing stimulus, say, TFs, the probability of activation above a threshold is again given by (40) with k_a and k_d replaced by $k_a^{'}$ and $k_d^{'}$ (Eq. (9)) . Fig. 5 shows $p(x > x_{thr})$, $x_{thr} = 0.25$, versus S/K_s in a semi-logarithm plot for both slow and fast kinetics. In the fast case, a substantially steeper curve is obtained leading to enhanced signal discrimination, i.e., a more predictable response in a gradient of inducing signal. As shown by Cook et al [3], the signal discrimination ability increases with gene copy number. One can thus speculate that diploid organisms utilise stochastic expression kinetics, preferably fast, for signal discrimination and are susceptible to degraded signal discrimination due to a reduction of gene copy number in the haploid state. Mutations in the subset of genes which generate a response to signaling gradients in diploid organisms may be the cause of some HI syndromes associated with these systems. Fig.

6 shows the same plot as in Fig. 5 but now the TF's form bound complexes with $n = 6$. In Eq. (9), $[S]$ is replaced by $[S_n] = K[S]^n$ (Eq. (12)). One now finds that the slopes of the curves for slow and fast kinetics are similar. Thus as n, the number of TF's forming the bound complex, increases, the distinction between slow and fast gene expression kinetics, as regards their signal discrimination ability, becomes less pronounced.

4. Summary

In this paper, we have studied simple mathematical models to explore the possible origins of HI. In Section 2, we have considered a deterministic model in which a complex of n TFs binds at the appropriate region of DNA to initiate gene expression in eukaryotes. The concentration of proteins synthesized versus the total concentration of TFs (Fig. 3) is a sigmoid. Due to the S-shape of the curve, the protein level may fall below a threshold when one of the two genes synthesizing the TFs is mutated, resulting in a 50% reduction in the total concentration of TFs. The absence of required protein activity can give rise to HI. In Section 3, we have studied simple stochastic models of gene expression and shown that due to random fluctuations around the mean protein concentration in the steady state, the protein level may fall below the threshold even though it does not do so in the deterministic case. The variance, a measure of the spread around the mean protein level, is reduced with increasing gene copy number and faster expression kinetics. The variance increases if the rate constant j_p associated with protein synthesis is increased. In the case of one gene, we have further calculated the probability that the concentration of proteins exceeds a threshold in the absence as well as the presence of an inducing stimulus. In the latter case, faster gene expression kinetics give rise to a sharper response to changing stimulus concentrations. As shown by Cook et al [3], this is also true when the gene copy number is increased. Thus the signal discrimination ability of diploid organisms may be impaired in the haploid state. When the inducing stimulus is a bound complex of n TFs, the distinction between slow and fast gene expression kinetics becomes less with increasing n. To sum up, we have considered both deterministic as well as stochastic models of gene expression and indicated possible scenarios for the occurrence of HI.

ACKNOWLEDGEMENT

R. K. was supported by the Council of Scientific and Industrial Research, India under Sanction No. 9/15 (239) / 2002 - EMR - 1.

References

[1] J. G. Seidman and C. Seidman, J. Clin. Invest. **109**, 451(2002)

[2] R. A. Veitia, BioEssays **24**, 175 (2002)

[3] D. L. Cook, A. N. Gerber and S. J. Tapscott, Proc. Natl. Acad. Sci. **95**, 15641 (1998).

[4] H. H. McAdams and A. Arkin, Trends in Genetics **15**, 65 (1999).

[5] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. **98,** 8614 (2001)

[6] T. B. Kepler and T. C. Elston, Biophysical Journal **81**, 3116 (2001)

[7] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc. Natl. Acad. Sci. **99,** 12795 (2002)

[8] N. G. Van Kampen, Stochastic Processes in Physics and Chemistry (North Holland, Amsterdam 1981).